Appendix

Appendix A1.1 Study characteristics: Nelson, Cooper, & Gonzales, 2005 (randomized controlled trial)

Characteristic	Description
Study citation	Nelson, J. R., Benner, G. J, & Gonzalez, J. (2005). An investigation of the effects of a prereading intervention on the early literacy skills of children at risk of emotional disturbance and reading problems. *Journal of Emotional and Behavioral Disorders, 13*(1), 3–12.
Participants	Forty-two kindergarten students with behavior problems were randomly assigned to either the intervention (*Stepping Stones to Literacy*) or the comparison condition. Three students who were performing at or above average with respect to phonological awareness skills were removed from each condition. Therefore, the analysis included 36 students (18 students per condition). Most of the participants were male students (17 males and one female in each condition). Minority students were 44% of the intervention group and 34% of the comparison group. The percentages of students eligible for free/reduced price lunch were 72% and 44% in the intervention and comparison groups, respectively. One student in each condition was an English language learner.
Setting	The study took place in seven elementary schools in a medium-sized Midwestern city.
Intervention	Over a five-week period, intervention group students received *Stepping Stones to Literacy* as a supplement to the core curriculum (*Open Court Reading* and early literacy developmental activities designed by the classroom teachers). The *Stepping Stones to Literacy* program consisted of twenty-five 20-minute one-on-one daily tutoring lessons. According to reports by tutors and independent observers, the tutoring sessions were implemented with a high level of fidelity to the *Stepping Stones to Literacy* curriculum.
Comparison	Comparison group students received the core curriculum and no other supplemental instruction. The study indicated that no attempt was made to change any of the teachers' regular instructional practices in the classroom.
Primary outcomes and measurement	The primary outcome measures were the Comprehensive Test of Phonological Processing (CTOPP): Phonological Awareness subtest and the Dynamic Indicators of Basic Early Literacy Skills (DIBELS): Initial Sound Fluency, Phoneme Segmentation Fluency, Letter Naming Fluency, and Nonsense Words Fluency subtests (see Appendix A2 for more detailed descriptions of outcome measures).
Teacher training	Information on training of tutors was not reported in the study.

Appendix A1.2 Study characteristics: Nelson, Cooper, & Gonzales, 2005 (randomized controlled trial)

Characteristic	Description
Study citation	Nelson, J. R., Stage, S. A., Epstein, M. H., & Pierce, C. D. (2005). Effects of a prereading intervention on the literacy and social skills of children. *Exceptional Children, 72*(1), 29–45.
Participants	Participants were 84 kindergarten students (64 in the intervention group and 20 in the comparison group) from 27 classrooms. Students were randomly assigned to conditions.[1] All students had behavior problems, which were identified based on high scores on a measure developed by Walker, Severson, & Gates (1995; as cited in Nelson, Stage, Epstein, & Pierce, 2005) to indicate risk for behavioral disorders. The second criterion for participating in the study was a low score on the DIBELS Letter Naming Fluency subtest. The analysis sample included 47 students in the intervention group and 16 students in the comparison group.[2] For the analysis sample, the study reported that 75% of the participants were male students, and about 26% were ethnic minority students. In addition, about 44% of the sample qualified for the free/reduced lunch program.
Setting	The participating students attended 10 elementary schools in the Midwest.
Intervention	The intervention was implemented during tutoring sessions, which were a supplement to the regular curriculum used at the schools. According to reports by tutors and independent observers, the tutoring sessions were implemented with a high level of fidelity.
Comparison	No information was provided for the comparison group other than that this group did not receive *SSL* services. The study indicated that no attempt was made to change any of the teachers' regular instructional practices in the classroom.
Primary outcomes and measurement	Primary outcome measures included the Comprehensive Test of Phonological Processing (CTOPP): Phonological Awareness subtest, the Dynamic Indicators of Basic Early Literacy Skills: Letter Naming Fluency subtest, and the Woodcock Reading Mastery Test–Revised: Word Identification and Word Attack subtests (see Appendix A2 for more detailed descriptions of outcome measures).
Teacher training	Information on training of tutors was not reported in the study.

1. The WWC has requested and received from the study author additional information about the assignment process. According to the first study author, 20 students' identification numbers were randomly selected from the eligible sample and assigned to the comparison group. The remaining students were assigned to the intervention group.
2. In keeping with the Beginning Reading Protocol, the WWC examined pretest/baseline scores and standard deviations of the post-attrition sample. The groups were similar at baseline (based on the WWC beginning reading review's convention of a mean difference less than 0.50 SD) and the study was not downgraded due to attrition.

Appendix A2 Outcome measures in the alphabetics domain by construct

Characteristic	Description
Phonological awareness	
Comprehensive Test of Phonological Processing (CTOPP): Phonological Awareness	A norm-referenced assessment that provides an overall measure of a child's phonological awareness skills. The composite score, which is based on three subtests, was used for rating purposes. The Elision subtest includes 20 items that measure the extent to which a child can say a word and then say what is left after dropping out designated sounds. The Blending Words subtest includes 20 items that measure a child's skill in blending separately presented sounds together to form words. The Sound Matching subtest includes 20 items that measure a child's skill in matching sounds (as cited in Nelson, Benner, & Gonzalez, 2005 and Nelson, Stage, Epstein, & Pierce, 2005).
Dynamic Indicators of Basic Early Literacy Skills (DIBELS): Phoneme Segmentation Fluency subtest	This standardized test measures a child's ability to segment three- and four-phoneme words into their individual phonemes fluently. The child is presented with words orally and asked to produce verbally the individual phonemes for each word (as cited in Nelson, Benner, & Gonzalez, 2005).
DIBELS: Initial Sound Fluency subtest	This standardized test measures a child's ability to identify the initial sound in an orally presented word. The child is presented with four pictures and associated names and asked to identify (by pointing to or naming) the picture that starts with the same sound presented orally by the examiner (as cited in Nelson, Benner, & Gonzalez, 2005).
Letter knowledge	
DIBELS: Letter Naming Fluency subtest	This is a subtest of a standardized measure in which students are presented with a page of upper- and lower-case letters arranged in a random order and are asked to name as many letters as they can. The score is the number of letters named correctly in one minute (as cited in Nelson, Benner, & Gonzalez, 2005 and in Nelson, Stage, Epstein, & Pierce, 2005).
Phonics	
DIBELS: Nonsense Words Fluency subtest	This subtest measures a child's word reading ability, including letter-sound correspondence and the ability to blend letter sounds into words (as cited in Nelson, Benner, & Gonzalez, 2005).
Woodcock Reading Mastery Test–Revised (WRMT–R): Word Identification subtest	This is a subtest of the norm-referenced WRMT–R. It includes 51 items that test decoding skills. It requires the child to read aloud isolated real words that range in frequency and difficulty (as cited in Nelson, Stage, Epstein, & Pierce, 2005).
WRMT–R: Word Attack subtest	This is a subtest of the norm-referenced WRMT–R. It includes 106 items that measure the child's ability to decode nonsense words. Students are aware that the words are not real (as cited in Nelson, Stage, Epstein, & Pierce, 2005).

Appendix A3 Summary of study findings included in the rating for the alphabetics domain by construct[1]

Outcome measure	Study sample	Sample size (schools/ students)	Authors' findings from the study: Mean outcome (standard deviation[2]) *SSL* group	Authors' findings from the study: Mean outcome (standard deviation[2]) Comparison group	WWC calculations: Mean difference[3] (*SSL* – comparison)	WWC calculations: Effect size[4]	WWC calculations: Statistical significance[5] (at $\alpha = 0.05$)	WWC calculations: Improvement index[6]
Construct: Phonological awareness								
Nelson, Benner, & Gonzalez, 2005 (randomized controlled trial)[7]								
CTOPP: Phonological Awareness	Kindergarten (low-achievers with behavior problems)	7/36	98.24 (9.40)	90.90 (9.60)	7.34	0.76	Statistically significant	+28
DIBELS: Phoneme Segmentation Fluency	Kindergarten (low-achievers with behavior problems)	7/36	19.43 (8.10)	11.20 (14.60)	8.23	0.68	Statistically significant	+25
DIBELS: Initial Sound Fluency	Kindergarten (low-achievers with behavior problems)	7/36	21.31 (7.90)	11.30 (7.60)	10.01	1.26	Statistically significant	+40
Nelson, Stage, Epstein, & Pierce, 2005 (randomized controlled trial)[7]								
CTOPP: Phonological Awareness	Kindergarten (low-achievers with behavior problems)	10/63	96.10 (11.50)	90.40 (10.50)	5.70	0.50	Statistically significant	+19
Construct: Letter knowledge								
Nelson, Benner, & Gonzalez, 2005 (randomized controlled trial)[7]								
DIBELS: Letter Naming Fluency	Kindergarten (low-achievers with behavior problems)	7/36	25.18 (10.60)	19.90 (16.90)	5.28	0.37	Statistically significant	+14
Nelson, Stage, Epstein, & Pierce, 2005 (randomized controlled trial)[7]								
DIBELS: Letter Naming Fluency	Kindergarten (low-achievers with behavior problems)	10/63	37.70 (14.70)	22.00 (13.40)	15.70	1.08	Statistically significant	+36

(continued)

Appendix A3 Summary of study findings included in the rating for the alphabetics domain by construct *(continued)*

Outcome measure	Study sample	Sample size (schools/ students)	Authors' findings from the study: Mean outcome (standard deviation[2]) — *SSL* group	Authors' findings from the study: Mean outcome (standard deviation[2]) — Comparison group	WWC calculations: Mean difference[3] (*SSL* – comparison)	WWC calculations: Effect size[4]	WWC calculations: Statistical significance[5] (at $\alpha = 0.05$)	WWC calculations: Improvement index[6]
Construct: Phonics								
Nelson, Benner, & Gonzalez, 2005 (randomized controlled trial)[7]								
DIBELS: Nonsense Words Fluency	Kindergarten (low-achievers with behavior problems)	7/36	12.34 (10.00)	3.90 (7.30)	8.44	0.94	Statistically significant	+33
Nelson, Stage, Epstein, & Pierce, 2005 (randomized controlled trial)[7]								
WRMT–R: Word Identification	Kindergarten (low-achievers with behavior problems)	10/63	104.80 (10.50)	94.30 (8.40)	10.50	1.03	Statistically significant	+35
WRMT–R: Word Attack	Kindergarten (low-achievers with behavior problems)	10/63	105.30 (10.60)	96.2 (9.80)	9.10	0.86	Statistically significant	+31
Average[8] for alphabetics (Nelson, Benner, & Gonzales, 2005)						0.80	Statistically significant	+29
Average[8] for alphabetics (Nelson, Stage, Epstein, & Gonzales, 2005)						0.87	Statistically significant	+31
Domain average[8] for alphabetics across all studies						0.84	na	+30

na = not applicable

1. This appendix reports findings considered for the effectiveness rating and the average improvement indices.
2. The standard deviation across all students in each group shows how dispersed the participants' outcomes are: a smaller standard deviation on a given measure would indicate that participants had more similar outcomes.
3. Positive differences and effect sizes favor the intervention group; negative differences and effect sizes favor the comparison group. The *SSL* group mean equals the comparison group mean plus the mean difference between the groups. The computation of the mean difference took into account the pretest difference between the study groups
4. For an explanation of the effect size calculation, see Technical Details of WWC-Conducted Computations.
5. Statistical significance is the probability that the difference between groups is a result of chance rather than a real difference between the groups.
6. The improvement index represents the difference between the percentile rank of the average student in the intervention condition versus the percentile rank of the average student in the comparison condition. The improvement index can take on values between –50 and +50, with positive numbers denoting results favorable to the intervention group.
7. The level of statistical significance was reported by the study authors or, where necessary, calculated by the WWC to correct for clustering within classrooms or schools and for multiple comparisons. For an explanation about the clustering correction, see the WWC Tutorial on Mismatch. See Technical Details of WWC-Conducted Computations for the formulas the WWC used to calculate statistical significance. In the case of Nelson, Benner, & Gonzalez (2005) and Nelson, Stage, Epstein, & Pierce (2005), no corrections for clustering were needed. In addition, no corrections for multiple comparisons were needed as the study reported on level of statistical significance after Bonferonni corrections.
8. The WWC-computed average effect sizes for each study and for the domain across studies are simple averages rounded to two decimal places. The average improvement indices are calculated from the average effect sizes.

Appendix A4 *Stepping Stones to Literacy* rating for the alphabetics domain

The WWC rates an intervention's effects in a given outcome domain as positive, potentially positive, mixed, no discernible effects, potentially negative, or negative.[1]

For the outcome domain of alphabetics, the WWC rated *Stepping Stones to Literacy* as having positive effects. The other ratings (potentially positive effects, mixed effects, potentially negative effects, and negative effects) were not considered because *Stepping Stones to Literacy* was assigned the highest applicable rating.

Rating received
Positive effects: Strong evidence of a positive effect with no overriding contrary evidence.
• Criterion 1: Two or more studies showing statistically significant *positive* effects, at least one of which met WWC evidence standards for a strong design.
Met. Two studies of *Stepping Stones to Literacy* showed statistically significant positive effects. Both studies met the WWC evidence standards for a strong design.
and
• Criterion 2: No studies showing statistically significant or substantively important *negative* effects.
Met. No studies showed indeterminate or negative effects.

1. For rating purposes, the WWC considers the statistical significance of individual outcomes and the domain-level effect. The WWC also considers the size of the domain-level effect for ratings of potentially positive or potentially negative effects. See the WWC Intervention Rating Scheme for a complete description.

Appendix A5 Extent of evidence by domain

Outcome domain	Number of studies	Sample size		Extent of evidence[1]
		Schools	Students	
Alphabetics	2	17	120	Small
Fluency	0	0	0	na
Comprehension	0	0	0	na
General reading achievement	0	0	0	na

na = not applicable/not studied

1. A rating of "moderate to large" requires at least two studies and two schools across studies in one domain and a total sample size across studies of at least 350 students or 14 classrooms. Otherwise, the rating is "small."